Multimodal Transportation 4 (2025) 100186

Contents lists available at ScienceDirect

Multimodal Transportation

journal homepage: www.elsevier.com/locate/multra

Full Length Article

Experimental determination of factors causing crashes involving automated vehicles

Teshome Kumsa Kurse[a,*], Girma Gebresenbet[b], Geleta Fikadu Daba[a], Negasa Tesfaye Tefera[a]

[a] Mechanical Engineering Department, Adama Science and Technology University, P.O. Box 1888, Adama, Ethiopia
[b] Division of Automation and Logistics, Department of Energy and Technology, Swedish University of Agricultural Science, P.O. Box 7032, 750 07, Uppsala, Sweden

ARTICLE INFO

Keywords:
Advanced technology
AVS
Clustering
Crash
CA DMV

ABSTRACT

Emergence of technologies to replace human action is occurring in many sectors, with autonomous vehicles being a leading example. Autonomous vehicles do not require human interaction and instead employ various devices to perform essential operations. This paper assesses factors which cause autonomous vehicles to suffer crashes, using field data collected by the Californian Department of Motor Vehicles. Data on these highly automated vehicles (AVs) were clustered based on degree and direction of impact, and analyzed by coding in Excel and RStudio programming. A novel feature of the work is that all clustering, analysis, application of association rules, and determination of degrees of severity of crashes were done by RStudio programming and that the direction of autonomous vehicles impacts was identified based on field data. Our analysis reveals that weather conditions, maneuvering, road conditions, and lighting are major factors in autonomous vehicles crashes. Rear-end crash and minor scratches to autonomous vehicles are the most frequent forms of damage, based on the available data. This study underscores the critical need for enhanced sensor technologies and improved algorithms to better handle adverse weather conditions, complex maneuvers, and varying road and lighting conditions. By identifying the most frequent types of damage, such as rear-end crashes and minor scratches, this research provides valuable insights for manufacturers and policymakers aiming to improve the safety and reliability of autonomous vehicles. The findings can inform future design improvements and regulatory measures, ultimately contributing to the reduction of crash rates and the advancement of autonomous vehicle technology.

1. Introduction

The concept of automated vehicles (AV) dates back to the 16th century, when Leonardo da Vinci created the first self-driving vehicle, a modest, three-wheeled, self-propelled cart (Bucolo M. B. A., 2020). It was a mechanical device, with a set of forces for engine power, a pre-programmable control mechanism, and an automatic parking brake triggered electronically by wire (Fuller, 2024). Thus, the idea of autonomous vehicles predates the automobile itself (Fig. 1). In the 20th century, scientists world-wide started creating the prototypes for modern autonomous vehicles, which are becoming closer to reality owing to technological breakthroughs in areas like computer vision.

* Corresponding author.
E-mail addresses: Teshome.kumsa@astu.edu.et (T.K. Kurse), Girma.gebresenbet@slu.se (G. Gebresenbet), gelefeke@gmail.com (G.F. Daba), Negasa.tesfaye@astu.edu.et (N.T. Tefera).

https://doi.org/10.1016/j.multra.2024.100186
Received 10 July 2024; Received in revised form 18 September 2024; Accepted 30 September 2024

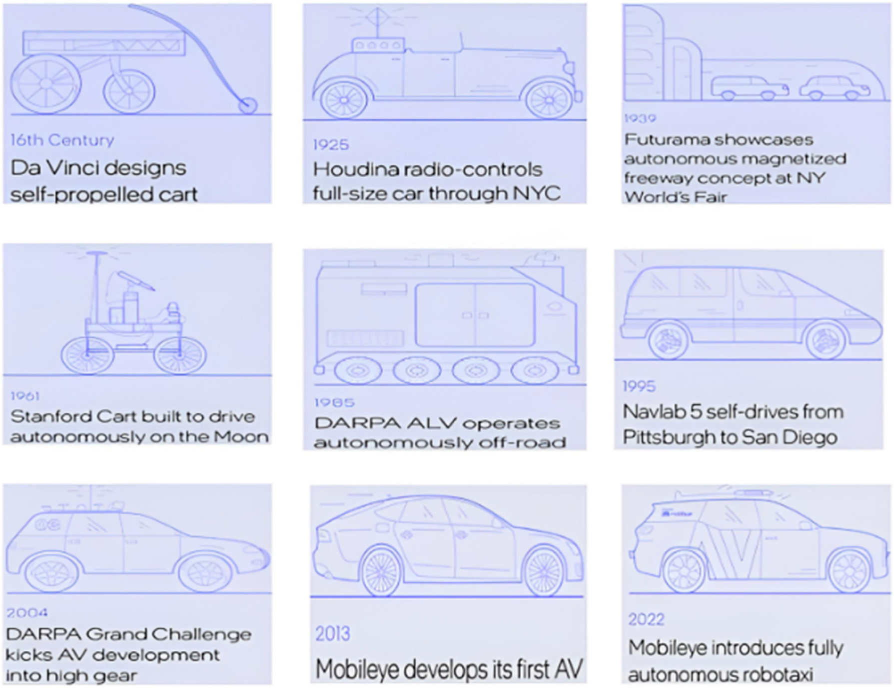


Fig. 1. Development history of autonomous vehicles (AVs) (Mobileye, 2023).

Table 1
Vehicle-controlling mechanisms in different development eras.

No.	Vehicle type/ year of invention	Controlling system
1.	Da Vinci's self-propelled cart, 1500.	High-tension springs propelled the cart and allowed for preprogrammed steering that directed it along a predetermined path.
2.	Whitehead Torpedo, 1868.	A torpedo that propelled itself underwater was a major changer for naval fleets worldwide. It could reach depths of hundreds of meters and maintained depth owing to a pressurization device (not explained).
3.	Mechanical Mike aircraft autopilot, 1933.	To ensure exact direction, gyroscopes tracked the plane's heading and communicated with the controls. Today, autonomous technology uses gyroscopes.
4.	Teetor Cruise Control, 1945 (applicable 1958).	One of the earliest cruise control systems, invented by an engineer who became irritated with the rocking motion he experienced when driving. It employs a mechanically driven throttle to calculate the speed of the car.
5.	Stanford cart, 1961.	Started with remote controlling and was later fitted with cameras and trained to recognize and follow a solid white line on the ground on its own.
6.	Tsukuba Mechanical Engineering, 1977.	Originally intended to be an autonomous passenger vehicle, it was designed to recognize traffic signs while accelerating beyond 20 miles per hour with the aid of two cameras mounted on the vehicle.
7.	VaMoRs, 1987.	Several cameras and sixty microprocessors were installed in a car to recognize items in front of and behind the vehicle, with the focus on pertinent objects.
8.	General Atomics MQ-1 Predator, 1995.	Equipped with car-adopted technologies, such as thermal imaging sensors that enable night-time driving and radar that can see through smoke or clouds.
9.	Tesla Autopilot, 2015.	With a single piece of software and a mix of cameras and radar, hand-free control is possible for driving on highways and motorways.

Fig. 1; shows recent steps in the progression from self-propelled cart to AVs. In each era of vehicle improvements, AV monitoring mechanisms have been updated (Table 1).

1.1. Automated vehicle crashes

Vehicle crashes are often an outcome of complex issues, but understanding the causes can contribute to accident prevention through different safety measures, regulations and training, to make the roads safer for everyone. A recent study proposes a new control method for highly automated vehicles that combines trajectory tracking and obstacle avoidance (Lin, 2020). This method uses predicted obstacle trajectories, lane-change replanning, and adaptive controllers to navigate safely and avoid crashes, as demonstrated through simulations. However, it has been reported that the primary problem with control systems is time delays, with an emphasis on real-time control of autonomous vehicles (Bucolo, 2019). On the other hand, (Xu, 2019), analyzed crash data from the California Department of Motor Vehicles to identify key factors influencing the severity and type of crashes involving connected and autonomous vehicles (CAVs), highlighting that driving mode, crash location, vehicle movements, and road conditions play significant roles.

Table 2
Raw data classification categories used in Excel.

Date	No. of vehicles involved	Accident details	Type of injury	Movement preceding crash	Crash type	Weather/lighting	Roadway surface	Roadway conditions	Weight of damage

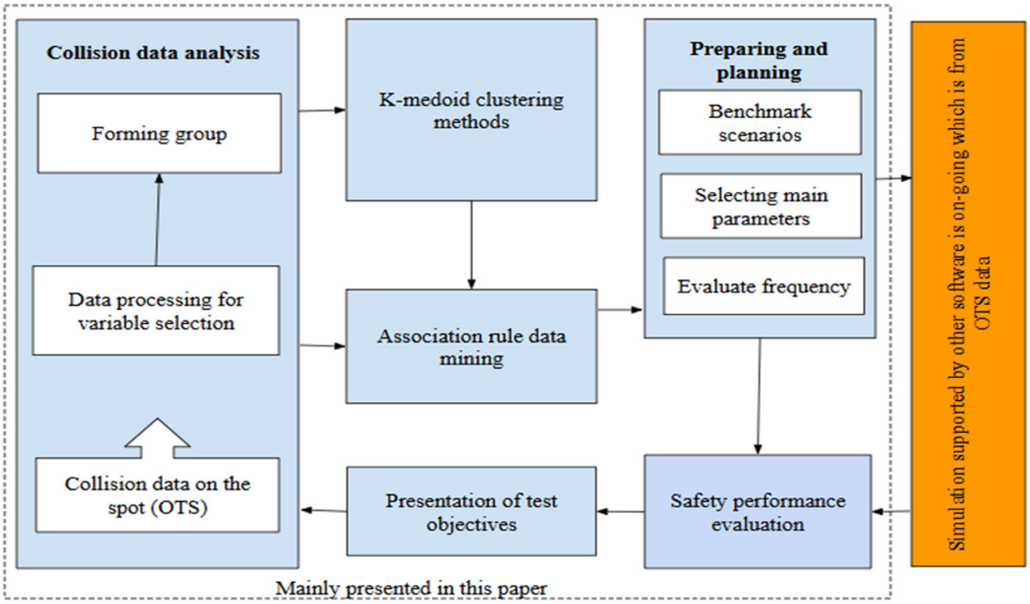


Fig. 2. Flowchart of the data handling process.

1.2. Motivations and target objectives

Research is ongoing on safety concerns relating to mixed vehicle populations, particularly traffic that consists of both driverless and driver-operated cars (Das, 2018). Even with the sophisticated on-board sensors, lidar, and radar used by manufacturing companies in current experiments, autonomous cars still have limitations. These include challenges in urban traffic situations, reliance on weather conditions, and unexpected behavior from other drivers. Expert surveys on the significance of road infrastructure, market preparedness, and the degree to which particular criteria influence the execution of designated automated driving functions on public roads have shown that the main issues for highly automated vehicles are complex urban environments, temporary work zones, and poor visibility brought on by bad weather conditions (fog conditions lead to deteriorated traffic flow characteristics, increasing crash risks) (Rahman, 2018; Gouda, 2021; Tengilimoglu, 2023; Fu, 2024). The main impacting elements are road surface conditions, road alignment, and illumination. However, the crash risks with autonomous vehicles and the kind of problems they may introduce to the current driving system are still unclear.

The aim of this study was to identify factors causing AVs to suffer crashes. This issue was investigated using field data collected by the Californian Department of Motor Vehicles (CA DMV) in field experiments (see Appendix A). Novel features of the work were that all clustering and identification of conditions leading to Automated vehicle crashes using data collected in the field were performed using the RStudio programming language and factors for Automated vehicle crashes are generally identified.

2. Methods

The CA DMV data used covered the period 2014 to August 2023 and encompassed crash type, position, place, AV type and status of the vehicle. These data are useful in determining how popular AVs are among the general public, how they affect traffic flow and congestion, and how safe and effective they are in various settings. The original dataset contained data on 639 crashes, but 11 of these lacked all necessary information and were rejected. Thus, data analysis was performed on 628 crashes, which were classified into different types in Excel spreadsheets (Table 2). The classified data were then converted to RStudio for more detailed assessment.

2.1. Data collection and processing style

The CA DMV database consists of on-the-spot (OTS) data, starting in 2014. Fig. 2 shows the data handling process used to investigate and identify safety factors that lead to accidents and injuries in road crashes involving autonomous vehicles.

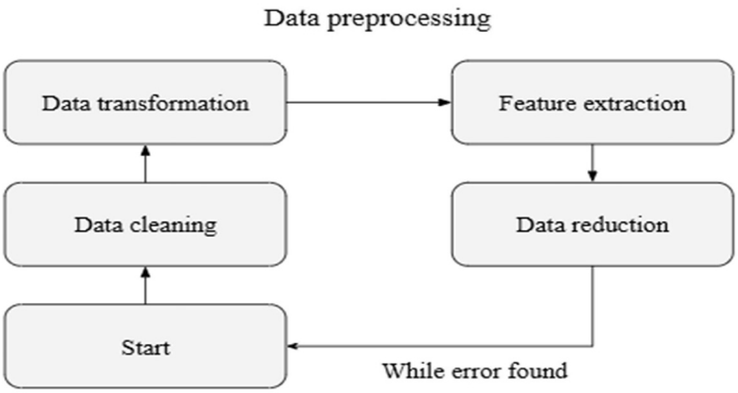


Fig. 3. Data preprocessing steps.

The flow of studying information in this research is methodically structured to ensure a comprehensive analysis of autonomous vehicle crashes. The process begins with Collision Data Analysis, where the initial step involves grouping the collected data based on various parameters such as date, time, location, weather, road conditions, lighting, and crash details. This grouping is essential for organizing the data after it has been collected and selected for further processing. Following this, K-Medoid Clustering is applied to cluster the data based on the degree and direction of impact, helping to identify patterns and common factors among similar crash incidents. Next, the study moves to Data Processing for Variable Selection. This involves Association Rule Mining, which analyzes the clustered data to uncover hidden patterns and correlations between different factors contributing to crashes. Association rules are crucial for identifying relationships that might not be evident through simple analysis. Additionally, the frequency of certain events or patterns is evaluated to understand their prevalence and impact on AV crashes.

In the Preparing and Planning Scenarios phase, benchmark scenarios are planned based on the analysis to test and evaluate the safety performance of AVs under different conditions. Key parameters are selected for further analysis and evaluation, ensuring that the most critical factors are considered. The Presentation of Test Objectives focuses on evaluating the safety performance of AVs. This involves assessing how well AVs handle various conditions and identifying areas for improvement. The process includes a feedback loop from “Data Processing for Variable Selection” back to “Collision Data Analysis,” indicating an iterative approach. This allows for continuous refinement and improvement of the analysis based on new data and insights. Finally, the study incorporates Simulation Outputs related to energy weight in Operational Traffic Systems (OTS), providing additional data to support the analysis and evaluation. By following this structured approach, the study systematically analyzes collision data, identifies key factors contributing to AV crashes, and evaluates the safety performance of AVs, ultimately aiming to improve their reliability and safety.

2.2. Data collection (exploration)

To gather all pertinent data pieces, including crashes that fulfilled the CA DMV portfolio requirements, data query and export tools were utilized.

2.3. Data preprocessing

Preprocessing to prepare the data for analysis and modeling is a crucial phase in data mining (Fig. 3). Applying proper preprocessing methods can prevent inaccurate findings caused by poor data quality

The raw data collection, classification, and preprocessing involved gathering detailed records of autonomous vehicle (AV) crashes reported to the Californian Department of Motor Vehicles (DMV). This data included various parameters such as the date and time of the crash, specific location details, weather conditions at the time of the crash, road conditions, visibility and lighting at the crash site, and detailed crash information including the degree and direction of impact, type of collision, and severity of damage. This comprehensive dataset provided on Table 2 below contributes a robust foundation for analyzing the factors contributing to AV crashes.

All types of intersections featured among the junction types involved in the 639 crash cases from the original OTS database. We decided to examine the data based on vehicle involvement in an accident, which meant that the data included crashes involving AVs either at junctions or in straight driving, and also included both conventional (human driven), and autonomous vehicles involved in crashes. From the compiled database, we found that the total number of vehicles involved in a crash ranged from 1 to 3, where the single vehicle caused an accident by contact with e.g., pedestrian, property, or cyclist. As the primary goal of the study was performance evaluation for autonomous cars, single-vehicle accidents were considered very relevant. Each sample greater or equal to two vehicles was connected to automated vehicles by assigning an ego and road user identity.

2.4. System of attribute selection and coding

Since vehicle performance evaluation needs data on specific characteristics, characteristics relevant to the purpose of the study were selected (see Table 2).

2.5. Parameters used

The given sample size was best suited for partitioning around medoids (PAM), which was selected based on the results, and duration of calculations. For error detection while information was transmitted, hamming distance was selected as a distance measuring tool. Silhouette analysis was found to be best approach to investigate separation of clusters. Table 3 shows sieved parameters for assessments of vehicle status, and conditions during accidents.

2.6. Specifying crash scenarios

As stated in the technical section (Borgelt, 2012; Luna, 2019; Rashmi, 2023), the clusters of damage types generated were further examined using a more powerful association rule mining system in order to identify the main factor in crashes.

2.6.1. Association rules/front itemset mining

This technique helps to uncover the relations between characters (Kaur M., 2016). It derives from market basket analysis (Ansari, 2019), which allow retailers to gain insight into commodities that are commonly purchased together, in order to optimize marketing campaigns and product shelving. The best example of this is the relationship between purchases of beer and crisps, where beer is the antecedent and crisps is the consequent. Multiple items can be included in a single itemset I. Using the association rules nomenclature, each sample is referred to as a transaction $(t_1, t_2, , t_n) \in T$, and each characteristic is referred to as an item $(i_1, i_2, , i_m) \in I$. An association rule may be expressed mathematically as follows: $X \to Y$ in which $X \subset I$, $Y \subset I$ and $XY = \emptyset$. Each rule is distinguished by its support and confidence values (Yusupova, 2019; Taherdoost, 2022):

$$Support\ (X) = \frac{|\{t \in T; X\ \ t\}|}{n} = P(X) \tag{1}$$

The support value for itemsets is the proportion of transactions t in the dataset that contains the itemset X. In the case of rules, support is defined as support of all items in the rule, i.e., $Supp\ (X \to Y) = Supp(X \cup Y) = P(X \wedge Y)$.

$$Conf\ (X \to Y) = \frac{Supp(X \cup Y)}{Supp(X)} = P(Y|X) \tag{2}$$

The conditional likelihood of a subsequent Y given the previous X is provided by confidence, which also assesses the strength of the rules. According to the alternative definition, it is the percentage of transactions that contain both X and Y. Assume that two rules with the antecedent and consequent flipped would have the same support value, in order to comprehend the differences between the two metrics. The most widely used implementation is the priori algorithm (Agrawal, 1993; Krishna, 2013), where locating association rules entails two steps:

1. Locate every item that occurs frequently and
2. Using the itemset that is collected, create an association rule.

A minimum confidence level and minimum support threshold are two requirements that the system must meet. An itemset is not common if its support criteria are smaller than the minimum level. In that scenario, every subset has to be uncommon and trimmable. Conversely, any subset of a frequent itemset has to be frequent. It is feasible to significantly reduce the number of alternative itemset configurations by using a straightforward algorithm using this idea iteratively. Generating rules from the frequently occurring itemset discovered in first step comprises the second step. As usual, the minimum confidence level comes into play. All non-empty subsets are constructed for each frequent itemset *I*.

In the event that the lowest confidence for this rule is given, then generate the rules (I_s) for each non-empty subset of *I*. Each rule also satisfies the minimum support as it is derived from sets of items that recur frequently. In this way, strong association rules can be found. The algorithm may produce millions of rules, depending on the intricacy of the data and how low the minimum support and confidence levels are set. Dedicated rule trimming and post-processing algorithms have been developed to find the most interesting rule. Previous findings indicate that confidence measures are not very useful in assessing how dependent the consequent is on the antecedent (Azevedo, 2007; Hahsler, 2015; Luna, 2018). In this work we used the lift metric, often known as 'interestingness':

$$Lift(X \to Y) = Lift(X \to Y) = \frac{Supp(X \cup Y)}{Supp(X) * supp(Y)} = \frac{P(X \wedge Y)}{P(X)P(Y)} \tag{3}$$

The incidence of X is negatively correlated with the occurrence of Y if the lift value is less than one, suggesting that the presence of one causes the absence of the other. When the result exceeds one, it indicates a positive correlation between X and Y, implying that the presence of one implies the presence of the other. If the lift is equal to one, then X and Y are independent (Kumar, 2016; Bao, 2021). Rules for interpretation can only be obtained at a minimum lift value greater than 1.

Table 3
Status of vehicles and conditions during accidents involving autonomous vehicles.

Category	Crash attributes		Count	Frequency
	Short name	Description		
Maximum injured	MaxInjd=Driver	Driver was injured by crash.	5	1 %
	MaxInjd=Injury	Random injury, no specifically explained.	75	12 %
	MaxInjd=None	No injuries, but vehicles were at risk.	50	8 %
	MaxInjd=Not explained	The crash was not explained fully.	33	5 %
	MaxInjd=Passenger	Passenger was injured in the crash.	7	1 %
	MaxInjd=Property	Property was damaged due to the accident.	451	72 %
Manoeuvre	Manvr=Passing other vehicles	The vehicle was passing next to another vehicle.	8	1 %
	Manvr=Backing	The vehicle was backing (reversing).	22	4 %
	Manvr=Changing lanes	The vehicle was changing lanes.	13	2 %
	Manvr=Entering traffic	The vehicle was entering traffic.	5	1 %
	Manvr=Making left turn	The vehicle was turning left.	37	6 %
	Manvr=Making right turn	The vehicle was turning right.	41	7 %
	Manvr=Merging	The vehicle was merging.	4	1 %
	Manvr=Parked	The vehicle was parked.	9	1 %
	Manvr=Parking manoeuvre	The vehicle was parking manoeuvre.	9	1 %
	Manvr=Proceeding straight	The vehicle was going straight ahead.	162	26 %
	Manvr=Slowing/Stopping	The vehicle was slowing to stop.	61	10 %
	Manvr=Stopped	The vehicle was stopped.	244	39 %
Type of crash	1stImpact=Broad side	First impact was on the vehicle's broad side.	36	6 %
	1stImpact=Head-on	First impact was head-on.	82	13 %
	1stImpact=Hit object	First impact was on an object.	27	4 %
	1stImpact=Overturned	First impact was while the vehicle overturned.	4	1 %
	1stImpact=Rear end	First impact was on the vehicle's rear end.	315	50 %
	1stImpact=Side swipe	First impact was while the vehicle was sideswiped .	118	19 %
	1stImpact=Not explained	First impact on the vehicle was not explained.	20	3 %
Roadway conditions	RdCond=Construction repair zone	The accident happened in a roadway construction/ repair zone.	28	5 %
	RdCond=No unusual conditions	When the accident happened there were no unusual conditions.	594	95 %
	RdCond=Other	When the accident happened, there was either flooded or holes, deep rut or obstruction on road way.	5	1 %
	RdCond=Reduced roadway width	When the accident happened, roadway width was reduced.	5	1 %
	RdCond=Not explained	Road conditions when the accident happened were not explained.	12	2 %
Road way surface	RdSurf=Dry	Dry road surface.	599	96 %
	RdSurf=Wet	Wet road surface.	27	4 %
Lighting	LightCond=DarkNSL	Darkness: no street lighting.	3	0 %
	LightCond=Dark Street lights	Darkness: street lights lit.	167	27 %
	LightCond=Daylight	Daylight present.	441	70 %
	LightCond=Dusk down	Dusk was about to fall.	13	2 %
Weather	Weather=Clear	At impact, weather conditions were clear.	571	91 %
	Weather=Cloudy	At impact, the weather was cloudy.	31	5 %
	Weather=Fog/Visibility	At impact, the weather was foggy.	5	1 %
	Weather=Raining	At impact, the weather was rainy.	19	3 %
Who involved in accident	1stInteract=Bicycle	The cyclist was first interacted with externally.	25	4 %
	1stInteract=None	No other party involved in the impact.	378	60 %
	1stInteract=Not explained	Not explained for first interaction.	135	22 %
	1stInteract=Other	Other was interacted with in the impact.	67	11 %
	1stInteract=Pedestrian	Pedestrian first interacted with in the impact.	6	1 %
	1stInteract=Scooter	Scooters first interacted with in the impact.	7	1 %
Vehicle motion status	1stVehMotion=Moving	At impact, the vehicle was moving.	358	57 %
	1stVehMotion=Stopped in traffic	At impact, the vehicle was stopped in traffic.	268	43 %

Table 4
Major factors in autonomous vehicle crashes and their implications.

No.	Factors in AV crash	Explanation for factor
1.	Communication error	AVs are increasingly being designed to communicate with each other and with infrastructure such as traffic lights. If there is a communication error, this could lead to autonomous vehicles making conflicting decisions and colliding with each other.
2.	Decision making errors	Even with accurate perception and prediction, highly automated vehicles can still make mistakes in decision making. This can happen due to incomplete or conflicting information, or due to limitations in the decision-making algorithms themselves.
3.	Environmental conditions	AVs can be affected by weather conditions, such as fog or rain, which make it difficult for their sensors to detect objects while moving. It implies that highly automated vehicles are weather conditions- dependent.
4.	Hardware failures	This can be experienced in conventional vehicles too. In autonomous vehicles hardware like sensors, cameras, and actuators may fail because of different external pressures. If any of these components fail, it could cause the AV to lose control, leading to crash.
5.	Human error	Highly automated vehicles are designed to be safe and smart, but human error can still contribute to crashes. As an example, a human driver may fail to follow the instructions of an AV, or may interfere with the AV's operation.
6.	Infrastructure	Highly automated vehicles can be affected by poor road conditions, physical infrastructure, and malfunctioning of traffic signals from digital infrastructures.
7.	Other drivers	Highly automated vehicles are designed to be safe and reliable, but they are still vulnerable to crashes caused by other drivers. For instance, a human driver may be inattentive, speeding, or operating a vehicle while under the influence of drugs or alcohol.
8.	Prediction errors	Highly automated vehicles need to predict the behaviour of other road users to make safe decisions. However, predicting human behaviour is complex and can lead to errors. For instance, an AV might assume a pedestrian will continue walking in a straight line when they suddenly decide to cross the road.
9.	Software errors (bugs and glitches) and sensor limitations	Vehicle operation relies on complex software systems. Glitches in these can cause the AV to behave an unexpected way which leads to a crash.
10.	Unexpected events	Even the best autonomous vehicles cannot anticipate every possible event on the road. As an example, a pedestrian or cyclist may step out in front of the AV suddenly. In these cases, the AV may not have enough time to react and avoid a crash.

2.6.2. Statistical parameters used

The minimum level of support and confidence is determined by the application and the intended outcome of the investigation. Ideally, rules with a lift value greater than one, high confidence, and strong support should be obtained. The focus of this work was on analyzing certain accident scenarios and features, which might be extremely unusual (Montella, 2012; Effati, 2015). Following testing with a range of values, an itemset that appears in <1 % of the sample is rejected when a minimum support of 0.01 is chosen. Decreasing the threshold lengthens the computation time and multiplies the rules that need to be understood. Selecting a higher support value might lead to omission of important cluster-related information. In the literature, there are several techniques for selecting a minimum confidence value. For example, a previous study on a powered two-wheeler (PTW) set a Conf = 0.1 threshold (Montella, 2011), which is lower than usual. Nonetheless, it is preferable to create rules in this task if there is a greater than 75 % chance of the consequent given the antecedent. Moreover, the results only take into account rules that have a lift larger than 1.25. The following process was used to eliminate duplicate rules in order to further reduce the quality of rules that were retrieved: the rule is deemed redundant if a more general rule with the same or higher lift already exists. In other words, a more specific rule is redundant again if it is only marginally or even less associated than a broader rule. A rule is broader if it has the same consequence, but one or more antecedents are eliminated. In formal terms, a rule $X \rightarrow Y$ is redundant if for $X' \subset X : lift(X' \rightarrow Y) \geq lift(X \rightarrow Y)$ (Hahsler, 2017)

3. Results

3.1. Main crash factors for vehicles

Most vehicle crashes happen due to different causes that are undisclosed or freely known, with the main categories being infrastructure, human error, and environmental conditions. Table 4 lists the top 10 factors that cause highly automated vehicles to collide and provides a short explanation of each. The subdivision into environmental factors, driver factors, and vehicle factors is based on reports from governmental and non-governmental research organizations, such as World Health Organization (WHO), the National Highway Traffic Administration (NHSTA), and the Insurance Institute of Highway Safety (IIHS).

3.1.1. Safety at road junctions

A crash database covering the years 2003 to 2013 was examined to obtain a picture of the scenarios in European Union junction accidents. The results showed that every third traffic-related accident involving human driven vehicles occurred at a junction, with about 43.9 % of fatalities and 43.2 % of severe injuries happening at crossroads. These percentages were also affected by junction type and by vehicle type. According to the OTS research by CARE (Janny Carson, 2023), people riding two-wheelers are killed more often at junctions than at any other part of the transportation systems. The most frequent crash types at signal-fitted crossroads are

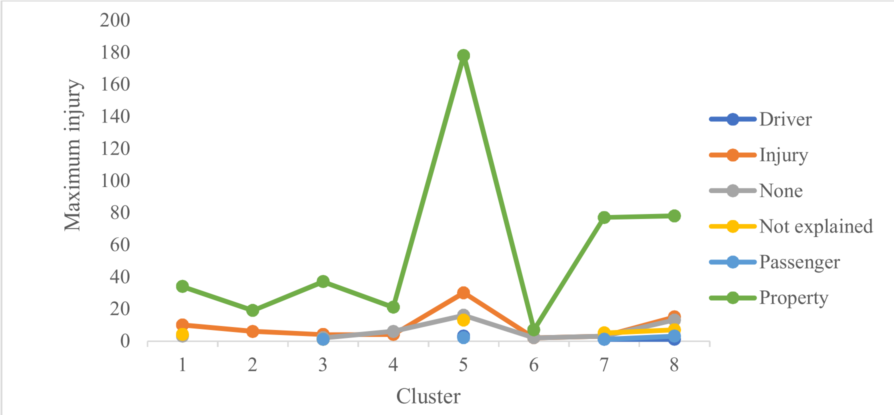


Fig. 4. Number of crashes with maximum injury per cluster.

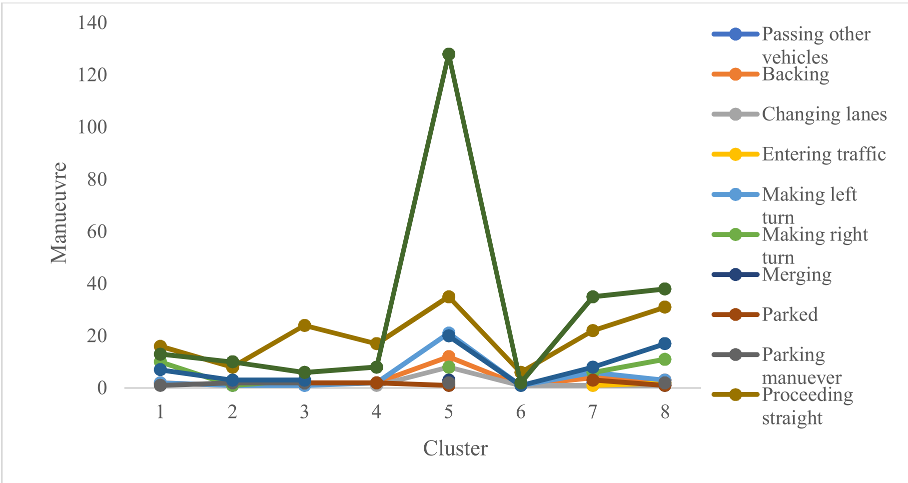


Fig. 5. Number of crashes involving maneuvering per cluster.

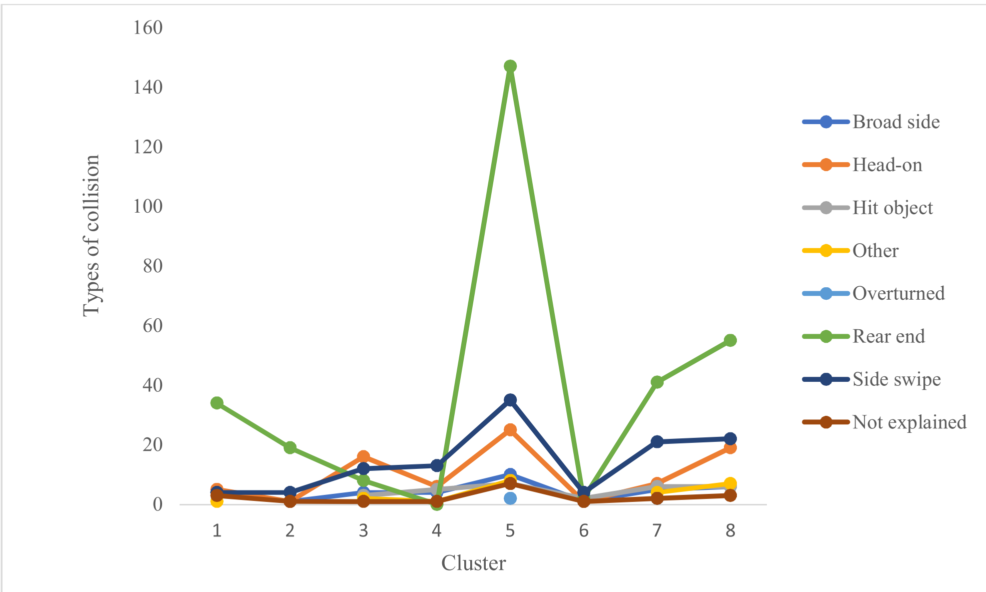


Fig. 6. Number of different types of crash per cluster.

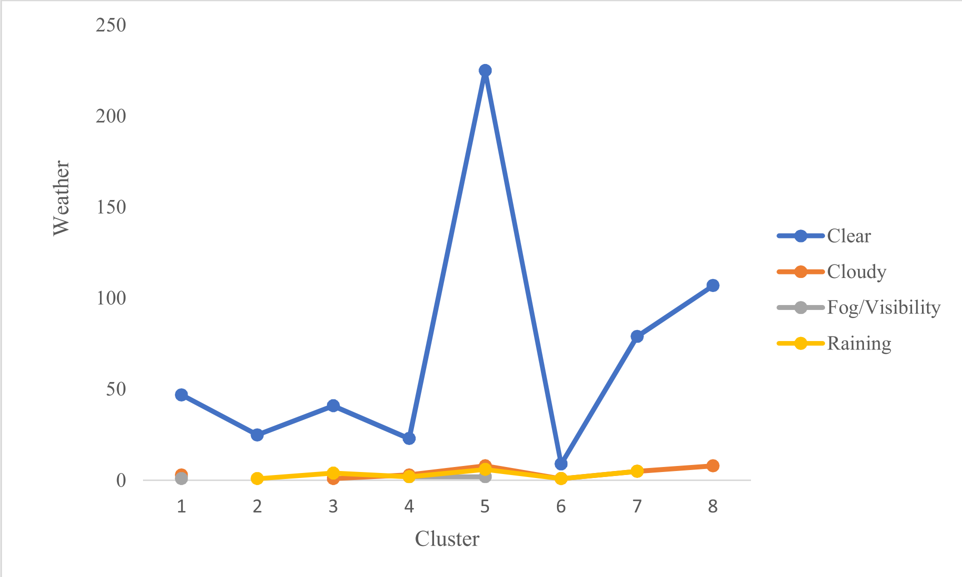


Fig. 7. Number of crashes involving weather conditions per cluster.

Table 5
Crash overcoming capacity of highly automated vehicles and human driven vehicles.

Human driven vehicles			Highly automated vehicles	
No.	Main causes of CV crashes	Explanations of causes	Main causes of AV crashes	Explanations of causes
1.	**Distracted driving**	This is the leading cause of all traffic accidents, including crashes at junctions. Drivers who are distracted by their phones, passengers, or other factors are more likely to miss traffic signals or other vehicles.	**Cyberattacks**	Highly automated vehicles are vulnerable to cyberattacks, and a successful attack could potentially cause an AV to make a mistake, such as running a red light.
2.	**Speeding**	Drivers who are speeding have less time to react to hazards, such as other vehicles entering the junction.	**Misinterpretation of traffic signals**	Autonomous vehicles may misinterpret traffic signals due to factors such as glare, dirt, or occlusion.
3.	**Impaired driving**	Drunk or impaired drivers are more prone to commit errors like running a red light or not yielding enough space to other traffic.	**Difficulty navigating complex intersections**	Autonomous vehicles may have difficulty navigating complex intersections with multiple lanes and traffic signals.
4.	**Aggressive driving**	This includes behaviour such as tailgating, cutting off other vehicles, and running red lights.	**Unpredictable behaviour of other road users**	Autonomous vehicles are designed to operate in a predictable environment. However, the unpredictable behaviour of other road users, such as pedestrians or cyclists, could potentially lead to crashes.
5.	**Right of way driving violations**	Drivers who fail to yield to vehicles with right of way, such as vehicles on the main road or vehicles making a left run, are also at risk of causing crashes at junctions.	–	–

reported to be head-on and rear-end crashes (Mohamed, 2017; Petrović, 2020). However, other studies suggest that this also depends on the number of lanes and traffic volume (Abdel-Aty, 2016; Jashami, 2023).

3.1.1.1. Difference in cause of crash at junctions for highly automated vehicles and human driven vehicles. When it comes to intersection safety, highly automated vehicles and conventional cars may differ in certain ways. Highly automated vehicles are equipped with sensors and software that allow them to perceive their surroundings and make decisions in a way that is not possible for human drivers. Table 5 provides other comparisons of highly automated vehicles and human driven vehicles.

3.1.2. Numerical investigations of crash frequency

Damage clustering and identification of the main reasons for AVs were performed, where each cluster had different numbers of samples. The clusters (C1-C8) were as follows: **C-1**: Most damage to the rear center and rear-right side of the AV, e.g., right rear tire, rear bumper and bumper bar, rear quarter panel, rear windshield, rear wheel and wheel wall, rear fascia, rear fender, rear passenger

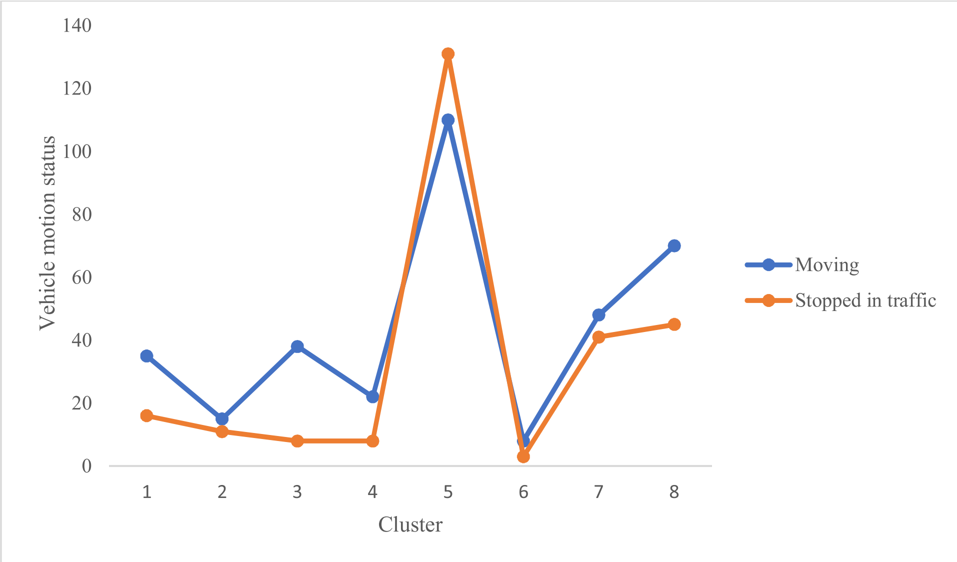


Fig. 8. Number of crashes involving different types of vehicle motion status per cluster.

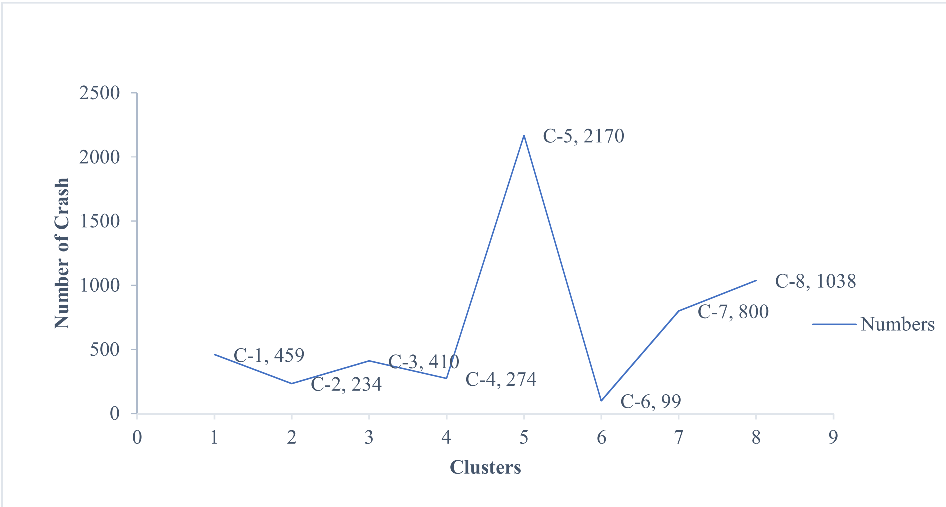


Fig. 9. Numerical explanations of vehicles in each cluster.

door, lower rear lamp, signal lamp and radar, radar mounting bracket and rear brake light. **C-2**: Damage to the rear-left side of the AV, e.g., left rear sensor, left corner sensor, left rear wheel, upper left tail lamp assembly, left rear door, vehicle left rear fascia, and rear driver side wheel molding. **C-3**: Damage to the front center and front right side of the AV, e.g., right brake assembly, front fender, right articulating radar casing, vehicle front center bumper, vehicle front, and front right tire. **C-4**: Damage to the front left side of the AV, e.g., driver side door, front driver side quarter panel, left bumper lidar sensor, driver side fender, driver side mirror grazed, front left wheel assembly, damage to the driver side radar assembly, and front left wheel assembly. **C-5:** Minor damage from front, rear, left or right direction with another vehicle or pedestrian. **C-6:** Major vehicle damage from front, rear, right and left direction. **C-7**: Vehicle and other property damage from different sides. **C-8:** Vehicle-related scratch, scrape, contact, and scuff marks on the AV.

Figs. 4-8 illustrate five selected parameters considered as factors in AV crashes, i.e., maximum injury (Fig. 4), maneuvering (Fig. 5), type of crash (Fig. 6), weather conditions (Fig. 7), and vehicle motion status (Fig. 8). In each diagram, the fifth cluster (C-5) is the largest, i.e., it covers more areas of vehicle damage, as indicated in Fig. 9.

Maneuvering status of the vehicle determined the position of damage. As Fig. 5 shows, a greater number of vehicle accidents occurred while the vehicle was stopping, either at a traffic light or parking at a parking area or on asphalt road. Stopping, proceeding

Table 6
Recommended solutions for autonomous vehicle crash risks.

No.	Solutions	Description of AV risk reduction at junctions
1.	Deploy updated traffic management systems	Traffic management systems can be used to coordinate the movement of highly automated vehicles and other vehicles on the road. This can help to reduce congestion and improve traffic flow, which can lead to fewer crashes.
2.	Develop better human-machine interfaces	Human-machine interfaces in highly automated vehicles should be designed to minimize the risk of human error. As an example, autonomous vehicles should have clear and concise displays that provide drivers with all of the information they need.
3.	Develop comprehensive safety standards (clear regulations)	We need to develop comprehensive safety standards for autonomous vehicles, including standards for testing and certification.
4.	Developing advanced prediction and more sophisticated decision-making algorithms	Future algorithms must incorporate a deeper understanding of human behavior and traffic patterns. Machine learning techniques can be used to train Autonomous vehicles on large datasets of real-world driving scenarios, while decision-making algorithms can be used to handle uncertainty and complex scenarios and implement failsafe mechanisms that ensure the AV takes a cautious approach when faced with ambiguous situations.
5.	Educate the public and drivers	We need to educate the public about autonomous vehicles and how they work. This will help to reduce the risk of human error and improve the acceptance of highly automated vehicles.
6.	Human Oversight	Even though highly automated vehicles are designed to be autonomous, they will still need some human oversight. This oversight could be provided by a human driver or by a remote operator.
7.	Improve AVs' ability to handle environmental factors	AVs should be equipped with sensors and software that can support them to perceive their surroundings and make safe decisions in a wide range of environmental conditions.
8.	Infrastructure improvement	If an updated system is going to become common in society in the future, infrastructure improvements like better traffic signals and dedicated lanes could reduce the risk of AV crashes.
9.	Invest in research and development	We need to continue to invest in research and development to improve AV technology and make it more reliable.
10.	Redundant systems	AVs should have redundant systems in place to minimize the risk of a crash due to a software error or sensor failure. As an example, highly automated vehicles could have multiple cameras and radar sensors.
11.	Through testing	Autonomous vehicles by nature need to be thoroughly tested in all kinds of conditions before they are deployed on public roads. This testing should include both simulated and real-world scenarios.

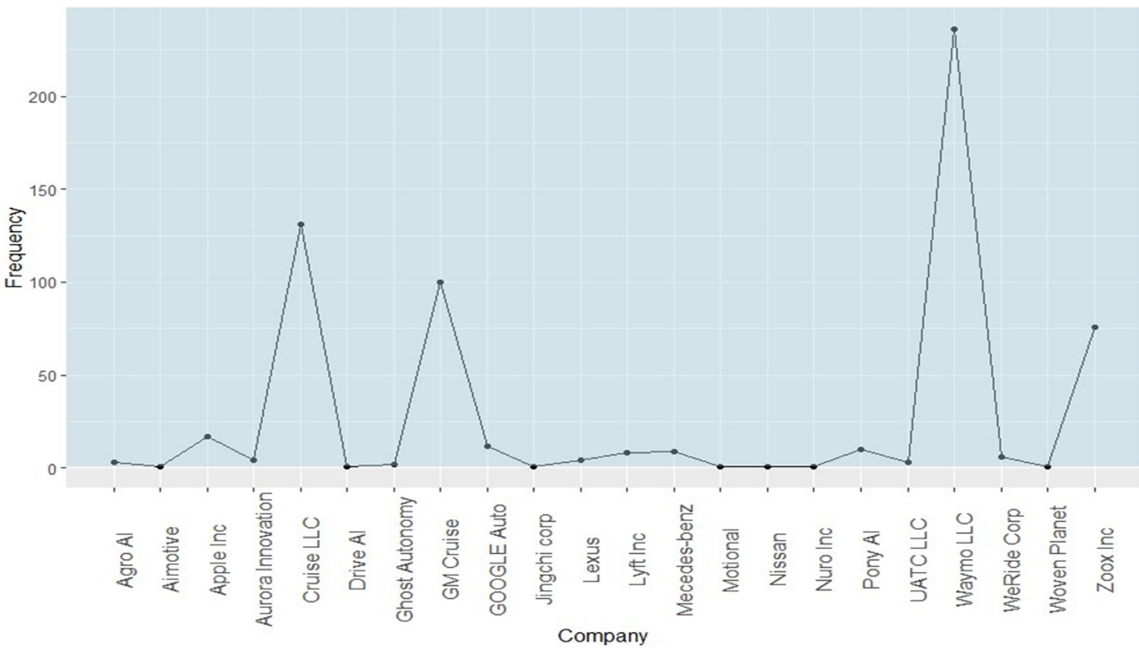


Fig. 10. Frequency of accidents involving autonomous vehicles from different manufacturers.

straight, slowing vehicle motion, turning, backing, changing lanes, and passing other vehicles were the main factors in accidents in this class.

Of the clustered data, most crashes involved cluster C-5, i.e., damage from front, rear, left or right direction with another vehicle or pedestrian (Fig. 9). The data indicates a significant variation in the number of crashes across different clusters. Cluster 5 (C-5) stands out with a very high number of crashes (2170), which could indicate a particular area or condition that leads to frequent accidents. This might require further investigation to understand the causes and implement safety measures. Cluster 6 (C-6), with the lowest number of crashes (99), could be an area or condition considered relatively safe. The remaining clusters (C-1 to C-4, C-7, and

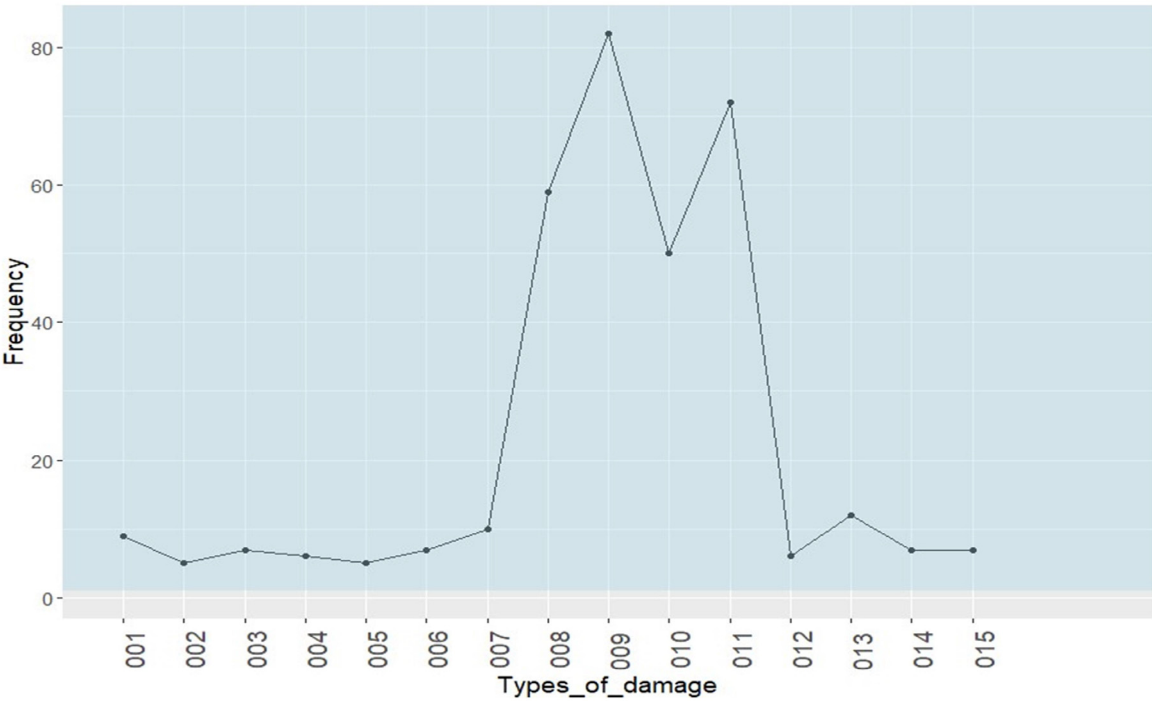


Fig. 11. Types of damage (coded 001–015) and frequency of each type:001 – Contact with vehicle front sensors 002 – Contact with rear sensors 003 – Damage to the front bumper 004 – Damage in the rear bumper and left rear sensor 005 – Front left radar damaging the sensor and its casing 006 – Minor cosmic damage to the vehicle 007 – Minor damage to the rear side of the vehicle 008 – Minor damage to the rear bumper 009 – Minor scratch on the vehicle 010 – Sustained vehicle damage 011 – Sustained minor damage 012 – Sustained minor damage to vehicle left rear bumper 013 – Sustained minor damage to vehicle rear bumper and hatch 014 – Sustained minor damage to the bumper 015 – Sustained minor damage to the side mirror.

C-8) have moderate crash numbers, suggesting varying degrees of safety or risk. This clustering helps in identifying patterns and areas of focus for improving vehicle safety and reducing crashes. Further analysis would be needed to understand the factors contributing to the high number of crashes in C-5 and the low number in C-6

At present, there are about 42 automated vehicles testing permission holders (about 36 with driver and six driverless). Among those, Waymo LLC holds permissions for most destinations (about 50 cities) for testing highly automated vehicles during both day and night. Our assessment using the RStudio algorithm to detect frequencies of vehicle crashes, based on data collected from CA DMV, showed that a greater number of accidents involved AVs operated by Waymo LLC (Fig. 10).

A review of all damage to AVs showed that minor damage, sustained damage, minor scratches, and minor damage to the rear bumper were the most frequent types of damage incurred in accidents (Fig. 11).

4. Discussion

Our assessment showed that highly automated vehicles are prone to contact with other vehicles, whether in autonomous or conventional mode, and manufacturers should pay more attention to overcoming this problem. Table 6 provides different recommendations to reduce the risk of crashes involving automated vehicles.

5. Conclusions

This study aimed to identify the primary factors contributing to crashes involving highly automated vehicles. Using data from the California Department of Motor Vehicle and advanced clustering techniques in RStudio, the analysis highlighted several critical factors influencing Automated vehicle crashes. These factors include communication errors, decision-making errors, environmental conditions, hardware failures, human errors, infrastructure issues, and errors in predicting other road user' behavior. The results of this assessment of causes of Automated vehicle crashes can help researchers and manufacturers reduce the risks. The results showed that most of the vehicles were damaged when stopping at the time of impact, due to e.g., traffic lights or during parking. An assessment of crash locations based on field data showed that a greater number of accidents, i.e., about 49 %, occurred on streets than at junctions, but that Automated vehicle safety at junctions was lower than that of human driven vehicles. Most of the crashes in the dataset were rear-end crashes, in conditions involving a dry roadway surface, daylight, and clear weather conditions.

The findings suggest that while highly automated vehicles are designed to enhance road safety, they remain susceptible to various challenges, especially those related their interaction with the environment and other road users. Improved communication systems, better decision-making algorithms, and enhanced sensor technologies are crucial for mitigating these risks. Additionally, addressing human factors and infrastructure improvements will be vital in reducing Automated vehicle crashes. Future research should focus on refining these areas to further enhance the safety and reliability of autonomous vehicles.

Data availability

The data that has been used is confidential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Teshome Kumsa Kurse: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Girma Gebresenbet:** Writing – review & editing, Writing – original draft, Funding acquisition, Data curation, Conceptualization. **Geleta Fikadu Daba:** Writing – review & editing, Funding acquisition, Conceptualization. **Negasa Tesfaye Tefera:** Writing – review & editing, Writing – original draft, Methodology, Data curation.

Acknowledgements

The authors would like to acknowledge Adama Science and Technology University for providing facilities particularly for support under the post graduate research fund scheme.

Appendix: A1

STATE OF CALIFORNIA
DMV
DEPARTMENT OF MOTOR VEHICLES
A Public Service Agency

REPORT OF TRAFFIC COLLISION INVOLVING AN AUTONOMOUS VEHICLE

DMV USE ONLY
AVT NUMBER
NAME

Instructions: Please print within the spaces and boxes on this form. If you need to provide additional information on a separate piece of paper(s) or you include a copy of any law enforcement agency report, please check the box to indicate "Additional Information Attached."

- Write **unk (for unknown)** or **none** in any space or box when you do not have the information on the other party involved.
- Give insurance information that is complete and which correctly and *fully* identifies the **company** that issued the insurance policy or surety bond, or whether there is a certificate of self-insurance.
- Place the National Association of Insurance Commissioners (NAIC) number for your Insurance or Surety Company in the boxes provided. The NAIC number should be located on the proof of insurance provided by you company or you can contact your insurer for that information.
- Identify any person involved in the accident (driver, passenger, bicyclist, pedestrian, etc) that you saw was injured or complained of bodily injury or know to be deceased.
- Record in the PROPERTY DAMAGE line any damage to telephone poles, fences, street signs, guard post, trees, livestock, dogs, buildings, parked vehicles, etc., including a description of the damage.
- Once you have completed this report, please mail to: Department of Motor Vehicles, Autonomous Vehicles Branch, 2415 1st Avenue, MS D405, Sacramento, CA 95818

SECTION 1 — MANUFACTURER'S INFORMATION

MANUFACTURER'S NAME		AVT NUMBER	
BUSINESS NAME		TELEPHONE NUMBER ( )	
STREET ADDRESS	CITY	STATE	ZIP CODE

SECTION 2 — ACCIDENT INFORMATION/VEHICLE 1

DATE OF ACCIDENT	TIME OF ACCIDENT ☐ AM ☐ PM	VEHICLE YEAR	MAKE	MODEL
LICENSE PLATE NUMBER	VEHICLE IDENTIFICATION NUMBER			STATE VEHICLE IS REGISTERED IN
ADDRESS/LOCATION OF ACCIDENT	CITY	COUNTY		STATE / ZIP CODE

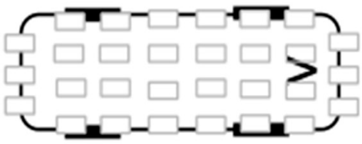
Vehicle was: ☐ Moving ☐ Stopped in Traffic
Involved in the Accident: ☐ Pedestrian ☐ Bicyclist ☐ Other ______
NUMBER OF VEHICLES INVOLVED

DRIVER'S FULL NAME (*FIRST, MIDDLE, LAST*)	DRIVER LICENSE NUMBER	STATE	DATE OF BIRTH
INSURANCE COMPANY NAME OR SURETY COMPANY AT TIME OF ACCIDENT	POLICY NUMBER		
COMPANY NAIC NUMBER	POLICY PERIOD FROM ______ TO ______		

Describe Vehicle Damage

☐ **UNK** ☐ **NONE** ☐ **MINOR** ☐ **MOD** ☐ **MAJOR**

Shade in Damaged Area



OL 316 (REV. 7/2020) **WWW**

Print Clear Form

SECTION 3 — OTHER PARTY'S INFORMATION/VEHICLE 2

VEHICLE YEAR | MODEL

LICENSE PLATE NUMBER | VEHICLE IDENTIFICATION NUMBER | STATE VEHICLE IS REGISTERED IN

Vehicle was: ☐ Moving ☐ Stopped in Traffic **Involved in the Accident:** ☐ Pedestrian ☐ Bicyclist ☐ Other ______ | NUMBER OF VEHICLES INVOLVED

DRIVER'S FULL NAME (*FIRST, MIDDLE, LAST*) | DRIVER LICENSE NUMBER | STATE | DATE OF BIRTH

INSURANCE COMPANY NAME OR SURETY COMPANY AT TIME OF ACCIDENT | POLICY NUMBER

COMPANY NAIC NUMBER | POLICY PERIOD FROM ______ TO ______

☐ **Additional information attached.**

SECTION 4 — INJURY/DEATH, PROPERTY DAMAGE

NAME (*FIRST, MIDDLE, LAST*)

ADDRESS | CITY | STATE | ZIP CODE

CHECK ALL THAT APPLY ☐ Injured ☐ Deceased ☐ Driver ☐ Passenger ☐ Bicyclist ☐ Property

NAME (*FIRST, MIDDLE, LAST*)

ADDRESS | CITY | STATE | ZIP CODE

CHECK ALL THAT APPLY ☐ Injured ☐ Deceased ☐ Driver ☐ Passenger ☐ Bicyclist ☐ Property

PROPERTY DAMAGE

PROPERTY OWNER'S NAME | TELEPHONE NUMBER ( )

STREET ADDRESS | CITY | STATE | ZIP CODE

WITNESS NAME | TELEPHONE NUMBER ( )

STREET ADDRESS | CITY | STATE | ZIP CODE

WITNESS NAME | TELEPHONE NUMBER ( )

STREET ADDRESS | CITY | STATE | ZIP CODE

☐ **Additional information attached.**

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

☐ Autonomous Mode ☐ Conventional Mode

☐ **Additional information attached.**

Print | Clear Form

OL 316 (REV. 7/2020) WWW

ITEMS MARKED BELOW FOLLOWED BY AN ASTERISK (*) SHOULD BE EXPLAINED IN THE NARRATIVE

	WEATHER (MARK 1 to 2 ITEMS)	VEH 1	VEH 2	MOVEMENT PRECEDING COLLISION	VEH 1	VEH 2	OTHER ASSOCIATED FACTOR(s) (MARK ALL APPLICABLE)
	A. CLEAR			A. STOPPED			A. CVC SECTIONS VIOLATED
	B. CLOUDY			B. PROCEEDING STRAIGHT			CITED
	C. RAINING			C. RAN OFF ROAD			☐ YES
	D. SNOWING			D. MAKING RIGHT TURN			☐ NO
	E. FOG/VISIBILITY			E. MAKING LEFT TURN			
	F. OTHER			F. MAKING U TURN			B. VISION OBSCUREMENT ☐
	G. WIND			G. BACKING			C. INATTENTION* ☐
	LIGHTING			H. SLOWING/STOPPING			D. STOP & GO TRAFFIC ☐
	A. DAYLIGHT			I. PASSING OTHER VEHICLE			E. ENTERING/LEAVING RAMP ☐
	B. DUSK – DAWN			J. CHANGING LANES			F. PREVIOUS COLLISION ☐
	C. DARK – STREET LIGHTS			K. PARKING MANUEVER			G. UNFAMILIAR WITH ROAD ☐
	D. DARK – NO STREET LIGHTS			L. ENTERING TRAFFIC			H. DEFECTIVE WEH EQUIP
	E. DARK – STREET LIGHTS NOT FUNCTIONING*			M. OTHER UNSAFE TURNING			CITED ☐ YES
	ROADWAY SURFACE			N. XING INTO OPPOSING LANE			☐ NO
	A. DRY			O. PARKED			I. UNINVOLVED VEHICLE ☐
	B. WET			P. MERGING			J. OTHER* ☐
	C. SNOWY – ICY			Q. TRAVELING WRONG WAY			K. NONE APPARENT ☐
	D. SLIPPERY (MUDDY, OILY, ETC.)			R. OTHER*			L. RUNAWAY VEHICLE ☐
	ROADWAY CONDITIONS (MARK 1 TO 2 ITEMS)			**TYPE OF COLLISION**			
	A. HOLES, DEEP RUT*			A. HEAD-ON			
	B. LOOSE MATERIAL ON ROADWAY			B. SIDE SWIPE			
	C. OBSTRUCTION ON ROADWAY*			C. REAR END			
	D. CONSTRUCTION – REPAIR ZONE			D. BROADSIDE			
	E. REDUCED ROADWAY WIDTH			E. HIT OBJECT			
	F. FLOODED*			F. OVERTURNED			
	G. OTHER*			G. VEHICLE/PEDESTRIAN			
	H. NO UNUSUAL CONDITIONS			H. OTHER*			

SECTION 6 — CERTIFICATION

I certify (or declare) under penalty of perjury under the laws of the State of California that the foregoing is true and correct.

I further certify that I am the authorized Administrator of the program for the above named employer.

PROGRAM DIRECTOR/AUTHORIZED REPRESENTATIVE PRINTED NAME AND TITLE	TELEPHONE NUMBER ( )
SIGNATURE **X**	DATE SIGNED

DL 316 (REV. 7/2020) **WWW**

Print Clear Form

Appendix: A2

```
library('ggpmisc')
library("ggpubr")
library(tidyverse)
library(readr)
library(lubridate)
library(dplyr)
library(tibble)
library(scales)
library(qpcR)
library(stringr)
library(data.table)
library(ggplot2)
Compilation1 <- read_csv("E:/Path/Folder/New folder/data.csv")
View(data)
knitr::kable(head(data[,1:14]), "pipe")
str(data)
column_types <- c(
Vehicles = "factor",
Number_of_vehicles_involved = "factor",
Vehicle_motion_status = "factor",
Manufacturer_name = "factor",
Accident_details = "factor",
Types_of_Injury = "factor",
Movement_preceeding_crash = "factor",
Crash type = "factor",
Weather = "factor",
Lighting = "factor",
Roadway_surface = "factor",
Road_way_conditions = "factor",
Vehicle_damage_description = "factor",
Weight_of_vehicle_damage = "factor")
data_cleaned <- na.omit(data)
manufacturer_counts <- data$Manufacturer_name %>%
factor %>%
table %>%
data.frame(
Company = names(.),
Frequency = .) %>%
mutate(Manufacturer_name = data$'Manufacturer_name'[match(Company, names(.))]) %>%
dplyr::group_by(., Manufacturer_name)
knitr::kable(manufacturer_counts, format = "pipe")
```

References

Abdel-Aty M, L.C., 2016. Identification of Intersections' Crash Profiles Patterns. Final Report. University of Central Florida.

Agrawal R, I.T., 1993. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207–216.

Ansari, S.Z. (2019). *Market basket analysis: trend analysis of association rules in different time periods.* https://run.unl.pt/handle/10362/80955.

Azevedo P J, J.a., 2007. Comparing rule measures for predictive association rules. In: European Conference on Machine Learning. Elsevier, pp. 510–517.

Bao F, M.L., 2021. An Improved Evaluation Methodology For Mining Association rules. *Axioms*. MDPI, p. 17.

Borgelt, C., 2012. Frequent item set mining. In: Wiley Interdisciplinary reviews: Data Mining and Knowledge Discovery, pp. 437–456.

Bucolo M, B.A., 2019. Forward action to make time-delay systems positive-real or negative-imaginary. Systems and Control letters. Elsevier.

Bucolo M, B.A., 2020. Automation of the Leonardo da Vinci machines. Machines (MDPI) 8 (3), 53.

Das, P., 2018. Risk Analysis of Autonomous Vehicle and Its Safety Impact On Mixed Traffic Stream. Rowan University, Glassboro, USA.

Effati M, S.-N.A., 2015. A semantic-based classification and regression tree approach for modelling complex spatial rules in motor vehicle crashes domain. In: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Wiley Online Library, pp. 181–194.

Fu C, L.Z., 2024. Distance headway-based safety evaluation of emerging mixed traffic flow under snowy weather. Physica A: Statistical Mechanics and Its Applications.

Fuller, J. (2024). *Did da Vinci really sketch a primitive version of the car?* Retrieved from howstuffworks: https://auto.howstuffworks.com/da-vinci-car1.htm.

Gouda M, C.I.-B., 2021. Automated assessment of infrastructure preparedness for autonomous vehicles. Automation in Construction. Elsevier.

Hahsler M, G.B., 2017. Introduction to arules-mining association rules and frequent item sets. In: SIGKDD Explor. Citeseer, pp. 1–28.

Hahsler, M., 2015. A Probabilistic Comparison of Commonly Used Interest Measures For Association Rules. Southern Methodist University, United States.

Janny Carson, G.J., 2023. Reducing Road Deaths Among Powered Two Wheeler Users. European transport safety council, Brussels.

Jashami H, A.J., 2023. Contributing factors to right-turn crash severity at signalized intersections: an application of econometric modeling. International Journal of Transportation Science and Technology. Elsevier.

Kaur M, K.S., 2016. Market Basket Analysis: identify the changing trends of market data using association rule mining. Procedia Comput. Sci. 78–85.

Krishna K A, A.D., 2013. Mining association rules between sets of items in large databases. Int. J. Sci. Modern Eng. (IJISME) 2319–6386.

Kumar S, J.N., 2016. Rule power factor: a new interest measure in associative classification. In: Procedia Computer Science. Elsevier, pp. 12–18.

Lin F, W.K., 2020. Integrated avoid collision control of autonomous vehicle based on trajectory re-planning and V2V information interaction. In: Sensors. MDPI, p. 1079.

Luna J M, F.-V.P., 2019. Frequent itemset mining: a 25 years review. In: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, p. e1329.

Luna J M, O.M., 2018. Optimization of quality measures in association rule mining: an empirical study. Int. J. Comput. Intell. Syst. 59–78.

Mobileye. (2023, February 27). *The History of Autonomous Vehicles from Renaissance to Reality*. Retrieved from https://www.mobileye.com/blog/history-autonomous-vehicles-renaissance-to-reality/.

Mohamed S A, M.K.-H., 2017. Investigating factors affecting the occurrence and severity of rear-end crashes. In: Transportation Research Procedia. Elsevier, pp. 2098–2107.

Montella A, A.M., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Anal. Prevent. 58–72.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. In: Accident Analysis & Prevention. Elsevier, pp. 1451–1463.

Petrović Đ, M.R., 2020. Traffic accidents with autonomous vehicles: type of collisions, manoeuvres and errors of conventional vehicles' drivers. In: Transportation Research Procedia, pp. 161–168.

Rahman M S, A.-A.M., 2018. Understanding the highway safety benefits of different approaches of connected vehicles in reduced visibility conditions. Transp. Res. Rec. 91–101.

Rashmi, K. (2023). *Data Mining in E-commerce: frequent itemset mining, association rules, and Apriori algorithm explained.* Retrieved from shiksha: https://www.shiksha.com/online-courses/articles/data-mining-frequent-item-set-mining-association-rules-and-apriori-algorithm.

Taherdoost, H., 2022. Different types of data analysis; data analysis methods and techniques in research projects. Int. J. Acad. Res. Manage. 1–9.

Tengilimoglu O, C.O., 2023. Implications of automated vehicles for physical road environment: a comprehensive review. Transportation Research Part E: Logistics and Transportation Review. Elsevier.

Xu C, D.Z., 2019. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. J. Safety. Res. 41–47.

Yusupova N, S.O., 2019. Data analysis methods for support decision making at management of complex systems. In: 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019), pp. 273–278.